Association of two different repetitive DNA elements near immunoglobulin light chain genes

Ronald Wilson and Ursula Storb

Department of Microbiology and Immunology, University of Washington, Seattle, WA 98195, USA

Received 17 December 1982; Revised and Accepted 14 February 1983

ABSTRACT

Sequence studies of repetitive DNA elements approximately 6 kb 3' of the mouse immunoglobulin C_κ region gene show that the R element located there (Gebhard et al. (1982) J. Mol. Biol. 157, 453-471) is adjacent to a 500 base pair long element which shows 80% homology to the BAM5 element sequenced by Fanning (Nuc. Acids Res. (1982), 10, 5003-5013). Neither the BAM5 element nor the R element itself is surrounded by a direct repeat, but the composite element (BAM5 + R) is surrounded by a 15 base pair direct repeat (with one mismatch). Direct repeats, consisting of target site sequences that surround a repetitive DNA element, are thought to arise during the insertion of the element at that site. It therefore appears that the BAM5 and R elements interacted and inserted as a linked entity. The existence of other BAM5/R composites throughout the mouse λ chain locus indicates that BAM5-R cooperation is not a rare event.

INTRODUCTION

Current research on characterizing repetitive DNAs falls essentially into two categories. There are studies which characterize repetitive DNAs that have been isolated and studied primarily because they *are* repetitive (1-10). In the second category, repetitive sequences are studied because of their proximity to other genes (11-15). We have chosen this latter strategy with the hope of eventually determining whether repetitive DNAs influence neighboring genes.

Analysis of repetitive DNAs in a variety of eukaryotes has revealed the existence of several families of repetitive DNA elements. Each element within a given family displays strong homology to other members of the same family. The most well characterized family, the *Alu* family in humans, and *Alu* equivalent families in other mammals, consists of approximately 500,000 members in the human (16,17). *Alu* elements almost always are situated within short direct repeat sequences (16), a configuration similar to prokaryotic transposable elements (18). *Alu* elements also usually exhibit a poly A run at one end (16,17). These two features prompted the suggestion

(19,20) that Alu elements propagate by a mechanism involving an RNA intermediate.

Data presented in this paper add to the growing body of evidence indicating that repetitive DNA elements, in some respect, interact with one another. Near the mouse immunoglobulin kappa constant region gene, we found a structure consisting of a BAM5 element (7) and an R element (12) linked together. Sequence data suggest that these two elements were mobilized as a linked entity. There are also such composite elements throughout the λ locus. We also present the results of an extensive computer search for homology between the R element and Alu type sequences. The homology suggests that even elements as disparate as R and Alu may have originated from a common ancestral sequence.

MATERIALS AND METHODS

Southern blot analysis (21) of the various genomic clones enabled us to locate repetitive DNA sequences on restriction maps. The genomic clones used were Sp101, a germline C_K clone (22); MOPC21-F2, a germline V_{K167} clone (23); KA19-V11, a germline $V_{K}MCP\text{-}11$-like clone (Wilson and Storb, unpublished); KA9 and KA5, germline clones containing the $C_{\lambda 3}$ and $C_{\lambda 1}$ genes and sequences to the 5' and 3' direction, respectively (24). Repetitive DNA sequences contained on these genomic clones were localized on the restriction map by digesting the clones with various enzymes (New England BioLabs), and constructing maps from blot data obtained using nick translated mouse kidney DNA as a probe (25).

Five repetitive DNA containing restriction fragments were selected for futher study, and were subcloned in pBR322. These five plasmids were from repetitive DNA regions on the C_K and the V_{K167} clones (see Figure 1a, and b). In cross hybridization studies, the plasmid clones were used as probes on blots containing digests of the genomic clones. Various portions of pRE103 that were subcloned into M13mp7 for sequencing (see below) were also used as probes radiolabeled by primer extension. Primer extension labeling was performed as follows: a dodecameric primer from Collaborative Research (Cat. No. 18061) was annealed to the template. The primer was extended using the conditions for dideoxy chain termination sequencing. There was no cold dNTP chase. Extended primers were removed from unincorporated nucleotides by gel filtration in Sephadex G75. In addition to the $C_{\lambda 3}$-$C_{\lambda 1}$ clones (KH5 and KA9) used above, germline clones covering the remainder of the λ locus were used in blot experiments with the primer extended

templates. They (24) are KX23, and KZ11, clones containing $V_{\lambda 2}$ and extending 5' and 3' of it, respectively; KX52, a clone containing $V_{\lambda 1}$ and 5' sequences; KX27, a clone containing $C_{\lambda 2}$-$C_{\lambda 4}$ and 3' sequences; KX61, a clone containing sequence just 5' of $C_{\lambda 2}$-$C_{\lambda 4}$, and KX107, a clone extending from 9 kb to 24 kb 5' of $C_{\lambda 2}$. For details, see Figure 3.

Sequencing was performed using the dideoxy method in conjunction with the cloning/sequencing vector M13mp7 (26). Plasmid pRE103 was cleaved with Sau3A I and ligated into Bam H-I cut M13mp7, or with Alu I, Rsa I, or Hae III and ligated into Hinc II cut M13mp7. Plates were plaque-lifted, filters were hybridized with mouse kidney DNA, and positive clones were sequenced. Sequence was obtained from genomic clones KH5 (Fig. 1d) and MOPC21-F2 (Fig. 1b) in a similar manner, except pRE103 was used as the probe for the plaque lifts. Sequence data were analyzed by homology search programs written in this laboratory (Wilson, unpublished).

Dot blots were made as follows. Liver DNA and pRE103 DNA were digested with Alu I, and then purified by phenol extraction and ethanol precipitation. They were then dissolved in 5µl of a buffer of 10 mM tris-Cl pH 7.5, 1 mM EDTA, and varying amounts of sonicated E. coli DNA. The amount of E. coli DNA plus test DNA was constant in each sample applied. The boiled samples were applied to a suspended nitrocellulose filter, which was then air-dried, and baked at 65°C overnight. Subsequent handling of the filter was identical to that of a Southern blot.

RESULTS

Localization of Repetitive DNA Sequences on Germline C_K and V_{K167} Gene Clones.

Various restriction digests of Sp101, a germline C_K gene clone, and MOPC21-F2, a germline V_{K167} gene clone, were analyzed on Southern blots using nick translated mouse kidney DNA as a probe. Such a probe detects repetitive DNA exclusively, since it lacks sufficient quantities of unique sequences to effect their detection (25). On Fig. 1a and 1b the arrows over the restriction maps show repetitive DNA containing fragments. Certain repetitive DNA containing fragments were subcloned in pBR322; the clone designations are shown above the arrows.

Sequence Homologies between various Repetitive Elements on Five Genomic Clones.

The two repetitive DNA plasmid subclones from the C_K clone Sp101 (pRE101 and pRE103) and the three from the V_{K167} clone MOPC21-F2 (pES301, pES304, pES312) were used as probes in further Southern blot experiments. Five genomic clones containing C_K, V_{K167}, V_{K11} and $C_{\lambda 3}$-$C_{\lambda 1}$ genes were

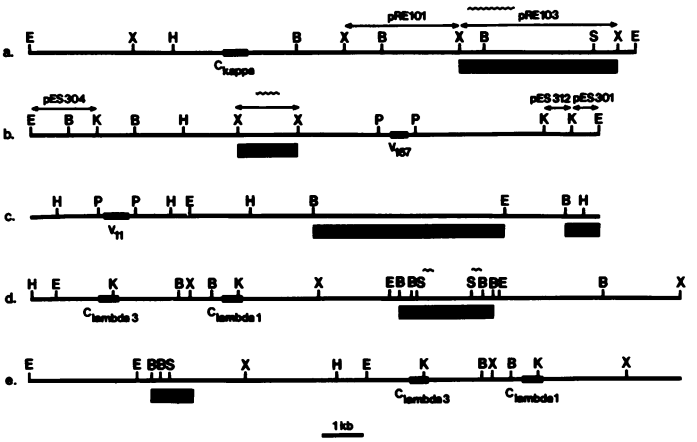


Figure 1. Restriction maps of five genomic clones; a. Sp101, a germline C_K clone, b. MOPC21-F2, a germline V_{K167} clone, c. KA19-V11, a germline V_KMPC11-like clone, d. KA9, a germline $C_{\lambda 3}$-$C_{\lambda 1}$ clone extending 5' of the genes, and e. KH5, a germline $C_{\lambda 3}$-$C_{\lambda 1}$ clone extending 3' of the genes. Sp101 (a) and MOPC21-F2 (b) were analyzed on Southern blots using nick translated mouse kidney DNA as a probe to detect repetitive DNA containing restriction fragments. Such fragments are indicated under the bidirectional arrows. Certain fragments were subcloned in pBR322, and are indicated by clone designations (e.g., pRE101, pRE103, etc.). Fragments exhibiting homology to pRE103 are shown over the thick black bars. E = Eco RI, X = Xba I, H = Hind III, B = Bam H-I, K = Kpn I, S = Sac I. All sites for a given enzyme are not necessarily shown. Regions which have been sequenced (figure 2) are shown by the wavy lines.

examined to determine whether they contained repetitive DNA sequences which would cross hybridize to the five plasmids. (These plasmid clones will obviously hybridize to the restriction fragments of the genomic clones from which they were derived.)

The only plasmid clone which cross hybridized to restriction fragments other than the ones from which it was derived was pRE103. These fragments are shown by the heavy bars in Fig. 1.

Sequence of Repetitive Elements of pRE103 and V_{K167}:

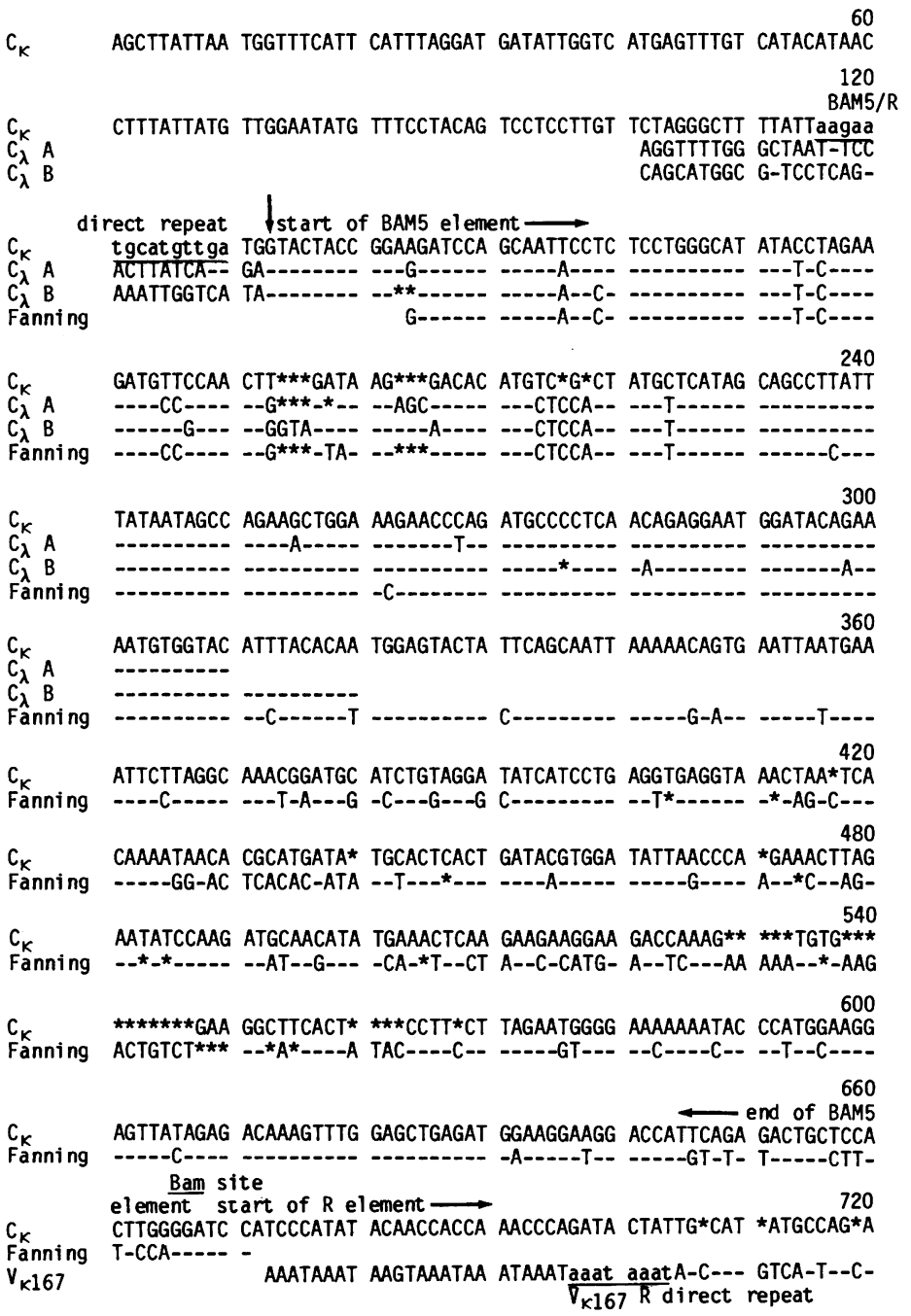
The repetitive elements were sequenced with M13 techniques (26). Repetitive DNA in pRE103 extended for approximately 1 kb (Fig. 2). A portion of this sequence has been described (12) and designated as a member of the R family of repetitive elements. The R element extends from positions 666 to 1153 in the sequence shown. The sequence from position 133 to 671 is similar to the BAM5 element described by Fanning (7). One striking feature of the sequence is that a 15 base direct repeat (with one

mismatch) borders the BAM5-R composite sequence (positions 116-130 and 1153-1167). The sequence order is 3' (relative to C_K gene transcription) - (direct repeat) - (BAM5) - (R) - (direct repeat) - (~6kb) - (C_K gene) - 5'.

Genomic V-region clones which hybridized to pRE103 were examined. Figure 1b shows that the V_{K167} clone MOPC21-F2 has sequences which cross hybridize to pRE103. The corresponding nucleic acid sequence is shown in Fig. 2. Its homology is only to the R element portion of the C_K sequence. Southern blot experiments indeed show that no BAM5 elements exist anywhere on the V_{K167} clone, nor on the KA19 (V_{K11}) clone. The R element consensus (12) extends from position 666 to 1173 as shown in Fig. 2. The R element found near V_{K167}, however, lacks homology up to position 706, and is therefore a truncated member of the family. In fact, it is similar to the truncated R5 sequence found 5' of the V gene in one of the rearranged kappa genes (T2) in myeloma T (12). In the sequence shown of the R element from V_{K167}, there is an eight bp direct repeat with one mismatch at the boundaries of the truncated element.

Presence of R and BAM5 Elements on Lambda Clones.

Since the pRE103 clone from the vicinity of C_K contained a BAM5/R composite element, and since both V_K clones tested lacked the BAM5 element, Southern blots were performed to determine whether R and BAM5 both were present in the pRE103-homologous sequences found in the immunoglobulin λ locus. Primer extended M13 templates which represented portions of the BAM5 and R sequences from pRE103 were used as probes. Figure 3 shows that R and BAM5 elements are found throughout the λ light chain locus. In all cases, except for a solo R element 3' of $V_{\lambda 2}$, the blot data imply that the two elements exist on the same or adjacent restricion fragments as composites. In some cases it was possible to map the orientation of R relative to BAM5 in a given composite, as shown near the $C_{\lambda 3}$-$C_{\lambda 1}$ genes. In other cases, both R and BAM5 could be assigned to the same restriction fragment. The map shows that the R/BAM5 composites 3' of $C_{\lambda 1}$ exist as a pair, in inverted repeat orientation, a finding substantiated by electromicroscopy (27). In contrast, the pair of composites 5' of the $V_{\lambda 1}$ gene do not correspond to the inverted repeat seen (27) in this region, indicating that the two composites are probably in direct repeat orientation. Direct repeat orientation of the composites 5' of $V_{\lambda 2}$ is also implied by the absence of any inverted repeat structure when $V_{\lambda 2}$ is analyzed by electron microscopy (27).


60
C_K AGCTTATTAA TGGTTTCATT CATTTAGGAT GATATTGGTC ATGAGTTTGT CATACATAAC
120
BAM5/R
C_K CTTTATTATG TTGGAATATG TTTCCTACAG TCCTCCTTGT TCTAGGGCTT TTATTaagaa
C_λ A AGGTTTTGG GCTAAT-TCC
C_λ B CAGCATGGC G-TCCTCAG-
direct repeat
start of BAM5 element
C_K tgcatgttga TGGTACTACC GGAAGATCCA GCAATTCCTC TCCTGGGCAT ATACCTAGAA
C_λ A ACTTATCA-- GA-------- ---G------ -----A---- ---------- ---T-C----
C_λ B AAATTGGTCA TA-------- --**------ -----A--C- ---------- ---T-C----
Fanning G------ -----A--C- ---------- ---T-C----
240
C_K GATGTTCCAA CTT***GATA AG***GACAC ATGTC*G*CT ATGCTCATAG CAGCCTTATT
C_λ A ----CC---- --G***-*-- --AGC----- ---CTCCA-- ---T------ ----------
C_λ B ------G--- --GGTA---- -----A---- ---CTCCA-- ---T------ ----------
Fanning ----CC---- --G***-TA- --***----- ---CTCCA-- ---T------ ------C---
300
C_K TATAATAGCC AGAAGCTGGA AAGAACCCAG ATGCCCCTCA ACAGAGGAAT GGATACAGAA
C_λ A ---------- ----A----- -------T-- ---------- ---------- ----------
C_λ B ---------- ---------- ---------- -----*---- -A-------- -------A--
Fanning ---------- ---------- -C-------- ---------- ---------- ----------
360
C_K AATGTGGTAC ATTTACACAA TGGAGTACTA TTCAGCAATT AAAAACAGTG AATTAATGAA
C_λ A ----------
C_λ B ---------- ----------
Fanning ---------- --C------T ---------- C--------- -----G-A-- -----T----
420
C_K ATTCTTAGGC AAACGGATGC ATCTGTAGGA TATCATCCTG AGGTGAGGTA AACTAA*TCA
Fanning ----C----- ---T-A---G -C---G---G C--------- --T*------ -*-AG-C---
480
C_K CAAAATAACA CGCATGATA* TGCACTCACT GATACGTGGA TATTAACCCA *GAAACTTAG
Fanning -----GG-AC TCACAC-ATA --T---*--- ----A----- -----G---- A--*C--AG-
540
C_K AATATCCAAG ATGCAACATA TGAAACTCAA GAAGAAGGAA GACCAAAG** ***TGTG***
Fanning --*-*----- --AT--G--- -CA-*T--CT A--C-CATG- A--TC---AA AAA--*-AAG
600
C_K *******GAA GGCTTCACT* ***CCTT*CT TAGAATGGGG AAAAAAATAC CCATGGAAGG
Fanning ACTGTCT*** --*A*----A TAC----C-- -----GT--- --C----C-- --T--C----
660
end of BAM5
C_K AGTTATAGAG ACAAAGTTTG GAGCTGAGAT GGAAGGAAGG ACCATTCAGA GACTGCTCCA
Fanning -----C---- ---------- ---------- -A-----T-- -----GT-T- T-----CTT-
Bam site
element start of R element
720
C_K CTTGGGGATC CATCCCATAT ACAACCACCA AACCCAGATA CTATTG*CAT *ATGCCAG*A
Fanning T-CCA----- -
V_{K167} AAATAAAT AAGTAAATAA ATAAATaaat aaatA-C--- GTCA-T--C-
V_{K167} R direct repeat

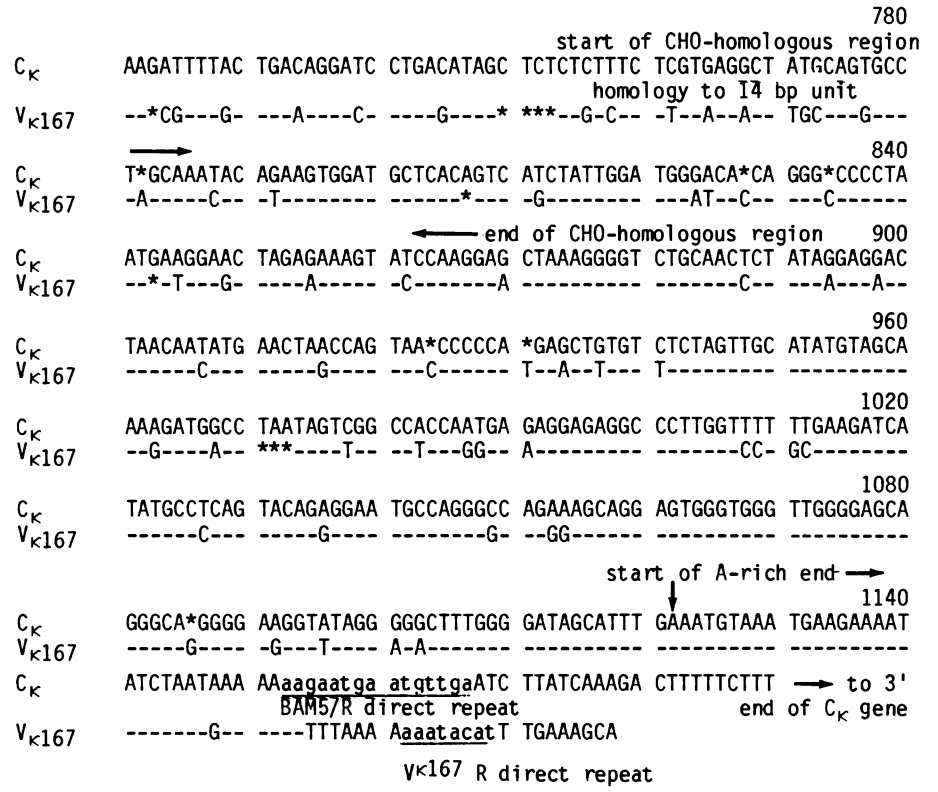


Figure 2. Sequence of the repetitive DNAs in pRE103 and the pRE103-homologous regions near V_{K167} and $C_{\lambda 3}$-$C_{\lambda 1}$. The repetitive DNA contained in pRE103 is compared to the homologous sequences found near the V_{K167} gene, positions 666 to 1153. It is also compared to the published BAM5 sequence (7) from positions 144 to 671. From positions 101 to 310 (and 320) the pRE103 sequence is compared to the left end of the two BAM5 sequences found 3' of $C_{\lambda 1}$. These two $C_{\lambda 1}$ associated elements are labeled $C_{\lambda}A$, and $C_{\lambda}B$. Sequence $C_{\lambda}B$ is from the BAM5/R composite nearer the $C_{\lambda 1}$ gene (Ritchie and Storb, unpublished). The direct repeats surrounding the BAM5/R composite and the V_{K167} associated R element are shown in lower case letters. Dashes indicate that the base in a given position is unchanged; an asterisk means that the base is missing.

Defining the ends of the BAM5 element.

The BAM5 element described by Fanning (7) is a sequence of a cloned 0.5 kb Bam H-I fragment obtained from digested mouse DNA. It is likely that this sequence does not include the entire sequence of the BAM5 family. Bam H-I must cleave within the consensus sequence, because the BAM5 family can be visualized as a band on ethidium bromide stained agarose gels of Bam H-I

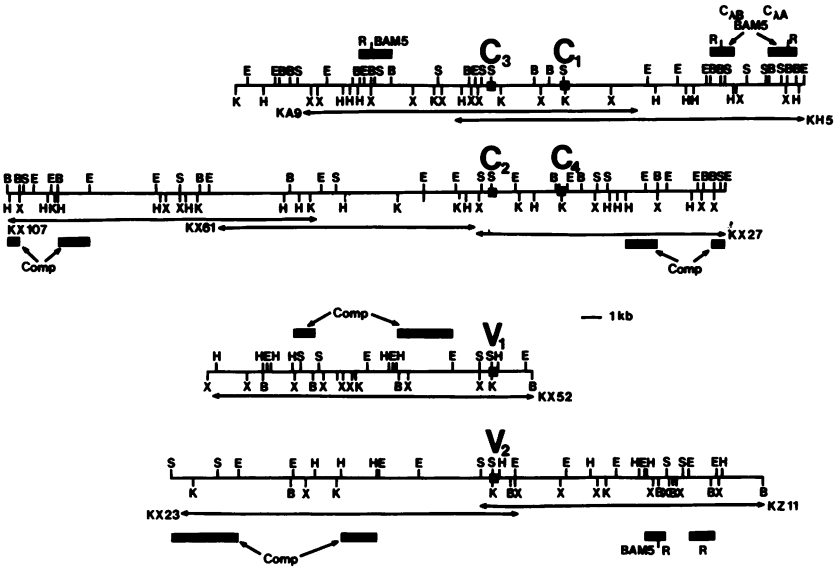


Figure 3. Location of BAM5/R composites near the λ locus. Genomic clones KA9 and KH5 were examined on Southern blots using M13 derived probes for BAM5 or for R sequences. The restriction fragments detected by the BAM5 probe are often adjacent to the restriction fragments detected by the R probe. Sequences $C_\lambda A$ and $C_\lambda B$ shown in Figure 2. Enzyme abbreviations are as in Figure 1. Although sequencing data establish the length of BAM5 to approximately 500bp, this figure depicts them as being larger. The bars indicate that we do not know their exact position within the restriction fragments in which they are found.

digested mouse DNA.

In Figure 2 we compare the sequence of the BAM5 near C_K and the analogous sequence from the $C_{\lambda 3}$-$C_{\lambda 1}$ associated BAM5 elements with Fanning's sequence. The consensus extends beyond the end of Fanning's sequence an additional 11 bp before the homologies discontinue (positions 143 to 132). The 15 bp direct repeat surrounding the C_λ associated BAM5/R composite element is an additional 2 bp beyond the end of the BAM5 consensus.

At the other end of the BAM5 element Fanning's sequence ends with a Bam site. This is the same Bam site marking the beginning of the R element in the C_K associated BAM5/R composite. (The R element consensus (12) also starts at this Bam site.) It is clear therefore that the BAM5 and the R elements in the composite are adjacent to one another with no intervening bases.

Table 1

	A		B	
	ng liver DNA	cpm probe bound	ng pRE103 DNA	cpm probe bound
R probe	500	271	128	1796
	250	181	64	1034
	125	75	32	634
	62.5	28	16	264
	31.25	13	8	143
	C		D	
BAM5 probe	125	37	32	503
	62.5	14	16	164
	31.25	4	8	73

Two dot blots were made (see Materials and Methods), each had liver DNA and pRE103 DNA on it, in the amounts shown in Table 1. One blot was hybridized to an M13 extended primer specific for R element sequences, the other to a BAM5 sequence specific probe. The dots were located by autoradiography and each subsequently cut out and counted in scintillant for 50 minutes. The cpm are shown. The four sets of data were then analyzed by linear regression. The results of the regression are as follows: A, 200 cpm probe bound required 341 ng of liver DNA; B, 200 cpm probe bound required 8 ng of pRE103; C, 50 cpm of probe bound required 162 ng of liver DNA; and D, 50 cpm of probe bound required 8 ng of pRE103 DNA. The pRE103 plasmid is approximately 8 kb long, while the R (or the BAM5 element) within it is 500 base pairs long. Therefore 8 ng of pRE103 corresponds to (500/8000) x 8 = 0.5 ng of either element. Considering data sets A and B, we know that 200 cpm of probe hybridized to pRE103 indicates 0.5 ng of element, thus in 341 ng of liver DNA, 200 cpm also detects 0.5 ng or 0.15% of the liver DNA. The mouse haploid genome is $3x10^9$ bases, so $(3x10^9)$ x $(0.0015) = 4.4x10^6/500 = 8800$ R elements per genome. A similar calculation for the BAM5 element results in 18,500 BAM5 elements per genome.

Dot Blot Analysis

To determine the number of R and BAM5 elements in the haploid genome, mouse liver DNA and pRE103 DNA were dotted onto nitrocellulose filters. It is known that 500 bases of each element are contained in the pRE103 plasmid, so it was determined how much liver DNA and how much pRE103 DNA would be required to hybridize to the same amount of probe. The data are presented in Table 1, and the computational details are in the legend. There are approximately 10000 copies of the R element, and 20000 copies of the BAM5 element per haploid genome.

DISCUSSION

BAM5 and R Association

Transposable elements in bacteria (18) generate direct repeats from the target sequences at their boundaries after insertion. It has been proposed

(19,20) that eukaryotic repetitive DNA sequences such as the Alu family propagate by RNA transcription of a given element, addition of adenine residues at the 3' end, reverse transcription creating the second strand, and subsequent reinsertion of a DNA copy of the transcript at another site. If the prokaryotic mechanism is applicable to eukaryotes, short DNA sequences at the target site are expected to become direct repeats formed at the ends of the element. Such a phenomenon in eukaryotes is strongly implied by recent data showing an Alu family green monkey equivalent element situated within satellite DNA between a direct repeat of nucleotides originating from the satellite DNA (28).

Several features of the sequences of R elements imply that they propagate in a manner similar to the way Alu elements are thought to propagate. R elements are surrounded by direct repeats, as are Alu elements, and both the Alu and R families have an adenine rich end, referred to here as the "right end".

Analysis of the sequence of the R element near V_{K167} suggests that it did arise from another R element. Although this R element is truncated at the "left end" (positions 666 to 705), the truncation could have originated by the following scheme. A full length R progenitor sequence is transcribed using a putative promoter, followed by removal of 40 bases at the left end. The now truncated element reinserts near V_{K167}, with the left end direct repeat (positions 707 to 714) adjacent to the left end of the shortened element.

At the right end (position 1150 to 1170) of the V_{K167} associated R element, the direct repeat (AAATACAT) is separated from the right end of the R consensus (12) by eight bases (ATTTAAAA). The Alu propagation model (19,20) predicts variation of the position of the right end direct repeat relative to the consensus, as there can obviously be variability in the number of A's added to the primary RNA transcript during the proposed polyadenylation step. (We believe subsequent mutation events altered what was previously a pure poly A run at the right end of these R elements.) The sequence data of Gebhard et al. (12) also show similar variation in the position of the right end direct repeats. In addition, their data show that the sequences which separate the right end direct repeat from the consensus are generally adenine rich.

Although the left end direct repeat of the V_{K167} R element directly abuts the element, this situation is not always the case, as evidenced by other R element sequence data (12). This positioning variation

suggests that sequences beyond the left end of the R consensus are also copied along with the element itself and subsequently co-inserted at the new site. The left end direct repeat would then be separated from the consensus by these additional bases. These extra bases seen in data of Gebhard et al. (12) contain no sequences derived from the BAM5 element. Although the positioning of direct repeats at both ends of the R elements can vary slightly relative to the consensus, the propagation scheme appears to be analogous to that of the Alu family (19,20,28). In short, R elements apparently give rise to other R elements.

BAM5/R composite elements also appear to give rise to other composites. The C_K associated composite is surrounded by a 15 base pair direct repeat (positions 116-130 and 1153-1167). The direct repeat positioning is similar to that found near single, non-composite elements. The left end direct repeat of the composite is two nucleotides from the consensus, and the right end direct repeat occurs after an A rich run. The junction of the right end of the BAM5 segment and the left end of the R segment is precise; there are no extra non-consensus bases (vicinity of position 666-671). Hence it appears that the composite inserted as a linked unit, presumably originating from another progenitor composite. Alternatively, the C_K associated BAM5/R composite could have arisen by an insertion of a BAM5 element near a pre-existing R element (or vice versa), but the sequence shows no evidence for this. There are no direct repeats surrounding either the R or the BAM5 segment alone. Such secondary event generated direct repeats could have been lost, but this appears unlikely since R and BAM5 and the associated direct repeats are so precisely joined. The sequence of the 7SL RNA of Hela cells is a clear example of a secondary insertion of one repetitive element near (or into) another (29). In this case, a middle repetitive sequence of 140 base pairs is found situated within an Alu family like sequence. This 140 base pair long element is itself surrounded by direct repeats.

Do BAM5 elements give rise to other BAM5 elements? Using available data, we would conclude that they do not, at least in the way that Alu and R elements seem to give rise to solo "progeny" elements. The genomic clones examined in this paper represent a total length of 164 kb of DNA. Within this DNA we find four solo R elements, several composites, and no solo BAM5 elements. Recent work (6-7) on the repetitive mouse MIF-1 element (2-7) implies (although there is no proof by DNA sequence), that the BAM5 element comprises one end of the approximately 5 kb long MIF-1 element. Since we find twice as many BAM5 elements as R elements in the genome, we suggest

that many of those not associated with R elements may be associated with MIF-1 and possibly other structures. (Gebhard et al. (12) have reported 10 times more R elements per genome than we have calculated (Table 1). Their estimate may have included more distantly related members of the family, because they used saturation hybridization techniques. We believe our data to be internally consistent.)

Is it possible that the BAM5/R composite is really a subelement of MIF-1? That does not seem to be the case for several reasons. 1) MIF-1 contains a 1.3 kb EcoRI fragment (6). No such fragment is found near the C_K associated BAM5/R composite. 2) The Kpn I site seen in MIF-1 is also absent from the C_K clone. 3) The size and overall restriction maps of MIF-1 (6) do not agree with the map of and the extent of repetitive DNA of the regions near the BAM5/R composite 3' of C_K which we have studied.

It may be possible that BAM5 elements can exist alone, i.e., surrounded by unique sequence DNA, but there are now two cases showing that BAM5 is a component of two very different higher order composite structures. If BAM5 is rarely or never situated alone, it may be unable to transpose by itself. If that were true, however, one might expect that it would become lost. It may be that BAM5 acts in some manner which enhances the chance for a transposable element associated with it to transpose.

The view now emerging is one in which composite elements, such as the BAM5/R composite, and MIF-1, are "superfamilies", or large elements constructed from smaller discreet families. This situation is similar to that found in sea urchins (30) where subelements are found scattered throughout repetitive DNA regions in a variety of different contexts.

Relation of Alu type sequences to R element sequences.

Jelinek et al. (31) pointed out that several seemingly unrelated sequences share homology with each other. These include the human Alu family, sequences at or near certain viral replication origins, RNA polymerase III transcription units near the globin locus, and sequences of some small RNAs of the Chinese hamster. Although the R sequence is not a mouse B1 element (the mouse equivalent of the human Alu element), it does share significant homology with this general class of sequences which were presented by Jelinek et al. The homology shown in Fig. 4 is between a portion of the mouse R element starting at position 752 and the Chinese hamster ovary type 2 Alu equivalent sequence (32). It has been proposed (19,20) that Alu family elements propagate through an RNA intermediate. This would be consistent with the poly A runs located at the end of Alu elements and located

```
          752 of Fig. 2
          ↓              14 bp homology region
R element CTCTCTTTCT *CGTGAGGCT ATGC*****A GTGCCTGCAA ATACAGAAGT GGATGCTCAC
Hamster       TGC--- T-CA----TC C--AGTTCA- T-C--A---- CC---**-*- --TG------

R element AGTCATCTAT    -57-    AGGGGTCTGC AACTCTATAG  -234 bases to poly A
Hamster   -ACA------    -20-    ACT---G--- -GA-A----T  - 28 bases to poly A
```

Figure 4. Partial homology of R element sequence to Chinese hamster type 2 Alu equivalent sequence. The R element sequence shown starts at position 752 of the C_K sequence shown in Figure 2. Dashes in the hamster sequence indicates that the base is the same; asterisks indicate a missing base.

within the putative insertion-event direct repeats. The homology shown in Fig. 4 is consistent with regard to poly A orientation; poly A runs in the R element and the CHO type sequence are at the same ends of the sequences shown, indicating that the homology shown has the same polarity. The homology shown in Fig. 4 also includes a run exhibiting strong homology to the 14 base consensus sequence (764-776) described by Jelinek et al. (31) which is located at or near viral replication origins. The R sequence starting at position 765 has preserved the invariant 5 base portion of the 14 base consensus. We conclude that at least portions of the R element and the Alu class of elements originated from a common ancestral sequence.

This portion of homologous bases in the R and Alu families could represent the now quite divergent relic of a subelement of the type described in sea urchins (30).

Conclusion

The data presented in this paper show that two elements of distinct repetitive DNA families evolved into a composite entity which has characteristics enabling it to propagate in the manner postulated for separate elements. Evidence is accumulating that some repetitive DNAs do have a regulatory function (29,33,34) although some may still exist solely for their own perpetuation (35). Further studies on repetitive DNA interaction may help clarify their biological role.

ACKNOWLEDGEMENTS

We thank Jeffrey Voss for determining the restriction map of the KA19-V11 (V_{K11}) clone, Dr. Erik Selsing for help in making the pRE101, and pRE103 clones, and for supplying the pES301, pES304, and pES312 clones. We also thank Kindred Ritchie for help in determining that the $C_\lambda A$

sequence (Figure 2) is 3' of the $C_\lambda B$ sequence, and Jim Miller for help with the BAM5/R composite map near $C_{\lambda 3}$-$C_{\lambda 1}$. We also would like to thank Ms. Ramona Finical for preparing the manuscript. Supported by NIH grants CA/AI 25754 and DE 02600 and NSF grant PCM78-13205.

REFERENCES

1. Houck, C.M., Rinehardt, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
2. Manuelidis, L. (1980) Nuc. Acids Res. 8, 3247-3258.
3. Cheng, S.-M. and Schildkraut, C.L. (1980) Nuc. Acids Res. 8, 4075-4090.
4. Heller, R. and Arnheim, N. (1980) Nuc. Acids Res. 8, 5031-5042.
5. Brown, S.D.M. and Dover, G. (1981) J. Mol. Biol. 150, 441-466.
6. Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F. and Bernardi, G. (1982) Proc. Natl. Acad. Sci. 79, 355-359.
7. Fanning, T.G. (1982) Nuc. Acids Res. 10, 5003-5013.
8. Anderson, D.M., Scheller, R.H., Posakony, J.W., McAllister, L.B., Trabert, S.G., Beall, C., Britten, R.J., and Davidson, E.H. (1981) J. Mol. Biol. 145, 5-28.
9. Donehower, L., Furlong, C., Gillespie, D. and Kurnit, D. (1980) Proc. Natl. Acad. Sci. 77, 2129-2133.
10. Krayev, A.S., Ksamerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A. and Georgiev, G.P. (1980) Nuc. Acids Res. 8, 1201-1215.
11. Shen, C.-K.J. and Maniatis, T. (1980) Cell 19, 379-391.
12. Gebhard, W., Meitinger, T., Höchtl, J. and Zachau, H.G. (1982) J. Mol. Biol. 157, 453-471.
13. Adams, J.W., Kaufman, R.E., Kretschmer, P.J., Harrison, M. and Neinhuis, A.W. (1980) Nuc. Acids Res., 8, 6113-6128.
14. Ryffel, G.U., Muellener, D.B., Wyler, T., Wahli, W. and Weber, R. (1981) Nature 291, 429-431.
15. Bell, G.I., Picket, R. and Rutter, W.J. (1980) Nuc. Acids Res. 8, 4091-4109.
16. Schmid, C.W. and Jelinek, W.R. (1982) Science 216, 1065-1070.
17. Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T. and Schmid, C.W. (1981) J. Mol. Biol. 151, 17-33.
18. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.
19. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T. and Gesteland, R.F. (1981) Cell 26, 11-17.
20. Jagadeeswaran, P., Forget, B.G. and Weissman, S.M. (1981) Cell 26, 141-142.
21. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
22. Walfield, A.M., Storb, U., Selsing, E. and Zentgraf, H. (1980) Nuc. Acids Res. 8, 4689-4707.
23. Selsing, E. and Storb, U. (1981) Cell, 25, 47-58.
24. Miller, J., Bothwell, A. and Storb, U. (1981) Proc. Natl. Acad. Sci. 78, 3829-3833.
25. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979) Nature 282, 567-574.
26. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nuc. Acids Res. 9, 309-321.
27. Selsing, E., Miller, J., Wilson, R. and Storb, U. (1982) Proc. Natl. Acad. Sci. 79, 4681-4685.

28. Grimaldi, G. and Singer, M.F. (1982) Proc. Natl. Acad. Sci. 79, 1497-1500.
29. Ullu, E., Murphy, S. and Melli, M. (1982) Cell 29, 195-202.
30. Posakony, J.W., Scheller, R.H., Anderson, D.M., Britten, R.J. and Davidson, E.H. (1981) J. Mol. Biol. 149, 41-67.
31. Jelinek, W.R., Toomey, T.P., Leinward, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L., and Schmid, C.W. (1980) Proc. Natl. Acad. Sci. 77, 1398-1402.
32. Haynes, S.R. and Jelinek, W.R. (1981) Proc. Natl. Acad. Sci. 78, 6130-6134.
33. Zuker, C. and Lodish, H.F. (1981) Proc. Natl. Acad. Sci. 78, 5386-5390.
34. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.
35. Orgel, L.E. and Crick, F.H.C. (1980) Nature 284, 604-607.